iCORE ANNUAL REPORT 2006

iCORE CPE GRANT CPE9

DR. RICHARD SUTTON

REINFORCEMENT LEARNING AND ARTIFICIAL INTELLIGENCE

ICORE ANNUAL REPORT MARCH 31, 2006

EXECUTIVE SUMMARY

A vast number of problems of economic and scientific interest involve sequences of actions where the effects of one action influence the expected utility of subsequent actions. These *sequential decision problems* include such diverse applications as inventory management, the control of robots and industrial processes, playing backgammon, and planning under uncertainty, all of which are made more challenging because of their sequential and stochastic aspects. Many problems in robotics and artificial intelligence are also of this nature, as indeed are most of the decision-making and planning problems faced by people and animals in their daily lives. Reinforcement learning is a new body of theory and techniques for solving sequential decision problems, based on classical methods such as dynamic programming and inspired by animal learning theory, that enables larger and more diverse problems to be solved.

The objectives of the reinforcement learning and artificial intelligence (RLAI) research program are to create new methods for reinforcement learning that remove some of the limitations on their widespread application, and to develop reinforcement learning as a model of intelligence that could approach human abilities.

In this third year, the RLAI research team has grown to about 40 members, 24 of which are graduate students and, of those, 11 are recipients of major scholarships. The output of the research program has also grown, with 39 papers published or accepted for publication in archival venues during the reporting period. After a long search, a fourth principal investigator was hired at the end of the year. Csaba Szepesvari of the Hungarian Academy of Sciences will start in September 2006 as an assistant professor at the University of Alberta. Dr. Szepesvari is internationally recognized as one of the foremost young theorists in reinforcement learning. He will be a PI of this project and also of the Alberta Ingenuity Center for Machine Learning.

The primary focus of the research program has been on how intelligent machines represent their knowledge of the world. The key question is how to organize the knowledge such that it can be verified, learned, and used autonomously without continual tending by human experts. This project has pursued an unusual approach in which knowledge is expressed in terms the machines' sensors and actuators, thereby enabling it to be compared directly to experiential data. Substantial progress was made this year toward formalizing the core learning algorithms and, to a lesser extent, toward developing planning algorithms.

The project has developed software this year, RL-Glue, that has become a de-facto standard for benchmarking reinforcement learning systems. RL-Glue is used throughout the world for research and education, and has taken Alberta a significant step closer to being internationally recognized as a leader in reinforcement learning research.

RESEARCH PROGRAM OVERVIEW

The iCORE research program in reinforcement learning and artificial intelligence (RLAI) pursues an approach to artificial intelligence and other engineering problems in which they are formulated as Markov Decision Processes and approximately solved using Reinforcement Learning. RL is a new body of theory and techniques for solving MDPs that has been developed in the last 20 years primarily within the machine learning and optimal-control research communities. RL researchers have developed novel methods to approximate solutions to MDPs that are too large or too ill-defined for classical solution methods such as dynamic programming. For example, RL methods have obtained the best known solutions in such diverse automation applications as helicopter flying, elevator scheduling, playing backgammon, and resource-constrained scheduling. The objectives of the RLAI research program are to create new methods for reinforcement learning that remove some of the limitations on their widespread application and to develop reinforcement learning as a model of intelligence that could approach human abilities.

The iCORE research proposal for the RLAI project identified four focus areas for investigation, on each of which substantial progress was made this year.

- One proposed focus was on *predictive state representations*, a new set of ideas for modeling sequential decision problems that are not Markov, that is, for which an appropriate state representation is not available a priori but must be constructed from the stream of raw sensor data. Last year we introduced a new formulation of predictive state representations, *temporal-difference networks*, that enables the representation of compositional predictions, a qualitative increase in abstraction abilities.

  This year we focused on the learning process in predictive state representations. We found a major flaw in previous learning methods – that they were only sound for behavior policies that ignored all observations, as typically would *not* be the case. We repaired the flaw by introducing a new class of learning algorithms that were not limited in this way. In addition, we extended temporal-difference networks to work with options, as discussed below.

- A second proposed focus was on *temporal abstraction*, the extension of RL beyond the flat and low-level representations commonly used with MDPs to the more flexible, structured, and higher-level representations used by AI systems. The ability to represent knowledge about possible courses of action at a multiplicity of inter-related temporal scales would vastly increase the generality and range of application of RL methods. The theory of options is one simple way of formalizing temporal abstraction. Last year we began combining options with temporal-difference networks.

  This year we have taken several steps further toward integrating the options framework with temporal-difference networks. Simple temporal-difference networks make predictions at each time step that are conditional on the primitive action taken on that

time step. Our extension enables the predictions to be conditional on options—multi-step closed-loop ways of behaving—rather than only on primitive actions. The options framework itself had to be substantially generalized in several ways to permit this extension, including new learning algorithms for the combination of recognizers and off-policy learning.

- A third proposed focus was broadly concerned with approximation and generalization in RL. Approximation is required in all large-scale applications yet is incompletely understood in both theory and practice. Last year we released an "RL toolkit" to facilitate the use of approximation and other RL techniques in applications.

  The year we have made significant progress with both software and learning algorithms. We have extended the RL toolkit with RL-Glue, a new interface standard that has become a de-facto international standard for RL competitions. We have also created a prototype RL-Library, a proposed repository for standard RL agents and environments. Finally, we have developed new algorithms for learning with linear function approximation when complicated by off-policy training, recognizers, and incremental least-squares methods.

- The fourth proposed area of investigation was the demonstration of advances in robotics applications. Last year we established a dedicated robotics laboratory and acquired a variety of new robots including a robot Segway.

  Our most striking demonstration this year was of robot *geo-caching* – the locating, usually by people, of a hidden cache through knowledge only of its GPS coordinates. Our attempt to have an autonomous Segway robot do geo-caching, while not entirely successful, nevertheless garnered substantial positive press attention both locally and internationally.

Some of these achievements are discussed in more detail in the following section.

Generally, the focus of the research program this year has continued to be on means by which an intelligent machine can represent knowledge of its environment or problem domain. Knowledge representation is widely recognized as critical to the performance of all AI systems, and critical to knowledge representation is expressing the knowledge in such a way that it can be verified, learned, and used without continual human monitoring. This project has pursued an unusual research strategy in which the knowledge is directly related to the machine's sensors and actuators. Temporal-difference networks with options are used to abstract from these low-level signals to the higher-level concepts required to solve very large problems, while maintaining verifiability and learnability. Over the next years we plan to extend these ideas and apply them to larger problems with the hope of addressing this long-standing challenge to all AI systems.

The RLAI web site can be found at http://rlai.net.

RESEARCH PROJECTS

These are some of the research projects ongoing this year.

Temporal Difference Networks

It has long been a goal for AI systems to be able to express their knowledge in terms of their interaction with the world. Connecting knowledge to experience provides a way of verifying and learning it autonomously. This ability may be essential to the practical development of large AI systems, and attaining it is a central objective of this research project. Last year we introduced a new framework and learning algorithm called "temporal-difference networks." Temporal-difference networks translate world knowledge into explicit predictive questions about future experience that can be compared with what actually happens. The major conceptual advance was putting the predictive questions in machine readable form (and not just their answers); prior work with prediction learning has taken the questions to be implicit, in the human designer's mind but not in the machine's. Representing them explicitly enables the set of questions being asked to be large and subject to autonomous elaboration.

Grounding knowledge in experience is challenging because knowledge is high-level and conceptual whereas experience is low-level and sensori-motor. Several levels and types of abstraction are required to bridge the gulf. The advanced temporal-difference networks that we began exploring at the end of last year combine two abstraction technologies: *predictive state representations* for abstracting over state, and the *options* framework for abstracting over time. This year we have taken this combination has been further developed in papers and computational experiments. Two papers on this subject have been submitted and accepted for publication at the prestigious Conference on Neural Information Processing Systems. One describes our initial extension of temporal-difference networks to include options, and the other describes a further development of the learning algorithm to a more general off-policy form as discussed in the next section.

Function Approximation, Off-policy Learning, and Recognizers

Off-policy learning is learning about one way of behaving (the target policy) while actually behaving in some other way (the behavior policy). The problem of off-policy learning is that classical bootstrapping RL methods such as Q-learning, TD(λ), and dynamic programming can become unstable during off-policy learning if function approximation is used. As both bootstrapping and function approximation are thought to be essential for large-scale applications, and off-policy learning is currently seen as necessary for learning temporally abstract system models, this instability is a key stumbling block to extending RL abilities. Moreover, if all three of these could be combined it would maximize the potential power of temporal-difference networks.

There are methods for off-policy learning using importance sampling that are known to be stable, but they are also known to be very slow to learn. As part of the work with temporal-difference networks, we have developed a new concept, that of a "recognizer," which achieves faster off-policy learning. A recognizer observes behavior and accepts it, or not, as something that it is learning about. It *recognizes* a portion of the behavior as, in effect, corresponding to the target policy. Recognizers are used to condition the predictions made by temporal-difference networks on the options taken. This year we have further developed the theory and practice of off-policy learning with recognizers and linear function approximation. We have shown both empirically and theoretically that importance sampling methods using recognizers have lower variance and are much better conditioned than previous importance-sampling methods for off-policy learning.

Incremental Least-squares Temporal-difference Learning

Approximate policy evaluation with linear function approximation is a commonly arising problem in reinforcement learning, usually solved using temporal difference (TD) algorithms. This year we introduced a new variant of linear TD learning, called *incremental least-squares TD learning*, or *iLSTD*. This method is more data efficient than conventional TD algorithms such as TD(0) and is more computationally efficient than non-incremental least-squares TD methods such as LSTD. In particular, we show that the per-time-step complexities of iLSTD and TD(0) are O(n), where n is the number of features, whereas that of LSTD is O(n^2). This difference can be decisive in modern applications of reinforcement learning where the use of a large number features has proven to be an effective solution strategy. In computational experiments, we have shown that iLSTD converges faster than TD(0) and almost as fast as LSTD.

Standardized Reinforcement Learning Software

RL Toolkit: Last year we released RL Toolkit 1.0 – a collection of software and guidelines to facilitate the development of RL research and applications. The RL Toolkit software has been placed in the public domain and is used within the RLAI group and by other researchers and students throughout the world.

RL-Glue: This year we have augmented the RL Toolkit with a new standard software interface for inter-connecting RL agents and environments. Our interface software is known as *RL-Glue* because it provides glue for interconnecting RL agents, environments and benchmark code. The software enables agents and environments written in different languages to be interfaced. RL-Glue has been a success in that it is now used throughout the world for research and is beginning to be used for education. RL-Glue was used in the first ever RL benchmarking event, held at the Neural Information Processing Systems Conference in December 2005. Researchers from more than 30 countries used RL-Glue to produce results reported at this event. RL-Glue has also been selected for use in the

first ever RL competition to be held at the 2006 International Conference on Machine Learning. These two conferences are the premier forums for RL research and RL-Glue appears to be becoming a de-facto international standard.

RL-Library: The *RL-Library* is meant to be a central site for storing and organizing RL code based on standards such as RL-Glue. Such sites exist for other branches of machine learning, and the need for one in RL has been felt for some time, but it has not been possible to create it without a standard such as RL-Glue. At the Neural Information Processing Systems Conference we achieved a consensus within RL community for the site to be created and hosted at the University of Alberta. A prototype RL-Library is up and available from the RLAI project's main web page, http://rlai.net.

Robot Geo-caching

We have started to examine the use of robots in unstructured outdoor environments using the task of geo-caching. Geo-caching involves locating a hidden cache through knowledge only of its GPS coordinates. A graduate-level project class using the Segway RMP has resulted in a successful demonstration. The robot was able to robustly navigate up to 200 meters and locate the visual target. We received local press (Edmonton Journal front page on December 14, follow-up story December 15) and national press (Daily Planet segment to air in January). The story was found repeated in places as far a Hungary (http://index.hu/tech/hardver/robotok/kato1223/).

Compact, Convex Upper Bound Iteration for Approximate POMDP Planning

Partially observable Markov decision processes (POMDPs) are an intuitive and general way to model sequential decision making problems under uncertainty. Unfortunately, even approximate planning in POMDPs is known to be hard, and developing heuristic planners that can deliver reasonable results in practice has proved to be a significant challenge. This year we have developed a new approach to approximate value-iteration for POMDP planning that is based on quadratic rather than piecewise linear function approximators. Specifically, we approximate the optimal value function by a convex upper bound composed of a fixed number of quadratics, and optimize it at each stage by semi-definite programming. We have demonstrated that this approach can achieve competitive approximation quality to current techniques while still maintaining a bounded size representation of the function approximator. Moreover, an upper bound on the optimal value function can be preserved if required. Overall, the technique requires computation time and space that is only linear in the number of iterations (horizon time).

OBJECTIVES FOR NEXT YEAR

The research team is expected to grow again next year. Csaba Szepesvari of the Hungarian Academy of Sciences and Mindmaker will start in September 2006 as a new principal investigator and as an assistant professor at the University of Alberta. Dr. Szepesvari is internationally recognized as one of the foremost young theorists in reinforcement learning. He will be a PI of the Alberta Ingenuity Center for Machine Learning as well. With his arrival the number of students and postdocs supervised is expected to increase.

A major focus in the upcoming year will continue to be on temporal-difference networks as a framework for grounding world knowledge in subjective experience. There are several further steps to be taken with regard to temporal-difference networks: the learning methods must be made more efficient, methods for automatically discovering questions and options must be found, and generally more experience and intuition have to be obtained regarding the power and limitations of temporal-difference networks. The near term objective is to develop more implemented examples to the point where a strong case can be made for the possibility and strength of grounding knowledge in experience, and then to prepare a journal article laying out the fundamental ideas of this approach.

We expect there to be a new major focus next year on planning methods with approximate, learned models of the environment. This will be an extension of ongoing work with iLSTD and with temporal-difference networks as predictive representations of world knowledge (environment models). We seek to demonstrate one system that can acquire an approximate, abstract model of the world's state and dynamics and then use that model to plan and act intelligently to maximize reward.

Next year we will release the first stable version of the RL-Library and encourage its widespread use in conjunction with RL-Glue. We will support the use of RL-Glue in the competition at the 2006 International Conference on Machine Learning, and we will also participate in that competition. RL-Glue will be extended so that it can be used over the internet, with agent and environment in different physical locations. A journal paper will be prepared documenting and explaining the structure of RL-Glue to the RL community, encouraging its widespread adoption. In addition, we will create a worldwide RL mailing list to be hosted at the University of Alberta.

Finally, this past year we began collaboration with Toyota Motor Corporation to explore the use of reinforcement learning technology for improving the fuel efficiency of gas-electric hybrid cars. We developed a hybrid car simulation with a spectrum of driving scenarios. In the coming year we would like to expand this collaboration by developing corresponding learning algorithms and demonstrating their potential in this application.

RESEARCH TEAM MEMBERS AND CONTRIBUTIONS

Faculty Team Members

Name	Role
Richard Sutton	Team leader and Principal investigator Reinforcement learning and artificial intelligence
Michael Bowling	Principal investigator Robotics, games, reinforcement learning
Dale Schuurmans	Principal investigator, CRC II Chair Probabilistic methods in artificial intelligence, machine learning
Vadim Bulitko	Associated faculty Decision-making in artificial intelligence
Mark Ring	Visiting professor, Chapman University Continual learning, neural networks, sequence learning. Dr. Ring participates in RLAI research by special arrangement as part of a collaboration with researchers at the Universities of Rutgers, Michigan, and Massachusetts
Katsunari Shibata	Visiting professor, Oita University, Feb-August 2005 Reinforcement learning, neural networks

Postdoctoral Fellows

Name	Topic/notes
Finnegan Southey	Machine learning, opponent modeling. Accepted position at Google
Li Cheng	Bayesian image modeling. Accepted research scientist position at the National ICT Australia
Yaakov Engel	Gaussian process reinforcement learning
Martin Zinkevich	Minimal regret algorithms
Mohammad Ghavamzadeh	Hierarchical reinforcement learning, Bayesian policy-gradient reinforcement learning
Tiberio Caetano	Structural pattern recognition; now a Postdoctoral Research Associate at the National ICT Australia

PhD Students

Name	Supervisor	Role / Topic	Awards / Special Info
Amir massoud Farahmand *	Sutton, Bowling, Jagersand	RL in robotics	
Ali Ghodsi	Schuurmans Brendan Frey	Tangent-corrected embedding	Ontario graduate scholarship University of Waterloo
Yuhong Guo	Schuurmans Russ Greiner	Maximum margin Bayesian networks	PhD Research Award
Jiayuan Huang	Schuurmans	Statistical language models	University of Waterloo, Internship at National ICT Australia (Jan-Jun 2006)
Feng Jiao	Schuurmans	Bayesian methods	University of Waterloo
Daniel Lizotte	Schuurmans Russ Greiner	Bayesian robot locomotion	NSERC, Killam/Steinhauer
Adam Milstein	Schuurmans	Action-respecting embedding	NSERC CGS-D University of Waterloo
David Silver *	Sutton Martin Mueller	RL in computer GO	iCORE, FS Chia, AIF
Brian Tanner *	Sutton	Temporal-difference networks	NSERC CGS-D, iCORE, AIF, Killam
Qin Wang	Schuurmans Dekang Lin	Large margin dependency parsing	iCORE
Tao Wang	Schuurmans/Bowling	Exploration/exploitation in RL	
Dana Wilkinson	Schuurmans	Action-respecting embedding	University of Waterloo
Linli Xu	Schuurmans	Structured predictors	University of Waterloo

MSc Students

Name	Supervisor	Role / Topic	Awards / Special Info
Nolan Bard	Bowling		
Alborz Geramifard *	Sutton/Bowling	iLSTD	
Michael Johanson	Bowling	Analyzing game play	Electronic arts funded
Armita Kaboli	Bowling/Petr Musilek	Robot localization	ECE student
Anna Koop *	Sutton	Predictive representations	AIF, iCORE
Mark Lee *	Sutton	RL-Glue and RL-Library	MSc student from Dec 1/05
Peter McCracken *	Bowling	Predictive state representations	NSERC, iCORE, Graduated Oct 2005. Awarded Outstanding MSc Dissertation award from Dept of Computing Science, U of A
James Neufeld *	Sutton/Bowling	Predictive representations in robots	

Cosmin Paduraru *	Sutton/Bulitko	Off-policy learning	iCORE
Eddie Rafols *	Sutton	Predictive representations	iCORE, NSERC, Walter H. Johns fellowship, Alberta Graduate
Adam White *	Sutton	RL-Glue and RL-Library	NSERC, iCORE, Walter H. Johns fellowship

* Directly supported by iCORE RLAI funding through Professor Sutton.

Other team members

Name	Role
Andrew Butcher	Graphics programming
Steve Fraser	Segway hardware (part-time)
Akiko Green	Administrative assistant (part-time)
Mark Lee	Programmer to Nov 30/05
Nelson Loyola	Aibo programming (part-time)
Jason Roberts	Segway programming
Christian Smith	Simulation programming (part-time)
Lori Troop	Program administrator (part-time)
Amanda Brewer	WISEST summer student
Priyanka Pareek	Open pages programming

COLLABORATIONS

Provincial	
Participants	Nature of Collaboration
Alberta Ingenuity Center for Machine Learning	Sutton, Schuurmans, and Bowling are among the seven principle investigators for this center at the U. of Alberta. AICML contributes substantially to the RLAI project, totaling approximately $800K for the reporting period

National	
Participants	Nature of Collaboration
Doina Precup, McGill University	Joint research on algorithms for off policy learning, temporal abstraction
Yoshua Bengio, McGill Hugh Chipman, U. Waterloo Christian Leger, U. Montréal William Welch, U. British Columbia Jim Ramsay, McGill Mu Zhu, U. Waterloo	MITACS grant "Statistical Learning of Complex Data with Complex Distributions" to Schuurmans, Sutton, and the listed collaborators $170K/year, RLAI portion ~$25K
Jonathan Schaeffer, Robert Holte, Duane Szafron, and Michael Buro University of Alberta	NSERC Strategic Grant, "Intelligent Agents for Interactive Entertainment" to Bowling, Schuurmans, and the listed collaborators $195K/year, RLAI portion ~$30K

International	
Participants	Nature of Collaboration
Satinder Singh, University of Michigan	Joint research with Sutton on predictive state representations and temporal-difference networks, conducted primarily by video-conference
Michael Littman, Rutgers University Satinder Singh, University of Michigan Andrew Barto, University of Massachusetts	DARPA grant "Intrinsically Motivated Reinforcement Learning" to Sutton with the listed participants. This grant funds Mark Ring's participation in RLAI $480K/year, RLAI portion ~$36K
E. James Kehoe, University of New South Wales, Australia	Joint research with Sutton and Neufeld on the relationship between reinforcement learning and learning in animals.

Industrial	
Participants	Nature of Collaboration
John Buchanan, Electronic Arts	With Bowling, on analyzing human play in computer games.
Michael James, Toyota Motor Corporation	With Sutton and Neufeld, exploring the use of reinforcement learning technology for improving the fuel efficiency of gas-electric hybrid cars.

The following table lists some of the collaborative activities of RLAI team members within the research community.

Collaborative activities in the research community		
Name	Conference/Journal/Activity	Role
Richard Sutton	Workshop on the "Frontiers of Reinforcement Learning" with Doina Precup from McGill	Co-organizer
Richard Sutton	Canadian Conference on Artificial Intelligence	Program committee
Richard Sutton	23rd International Conference on Machine Learning	Area chair
Michael Bowling	Workshop on "Multiagent Learning" at the 20th National Conference on Artificial Intelligence	Organizing committee
Michael Bowling	International Conference on Intelligent Robots and Systems	Exhibition co-chair
Michael Bowling	23rd International Conference on Machine Learning	Program committee
Michael Bowling	21st National Conference on Artificial Intelligence	Program committee
Michael Bowling	*Machine Learning* Journal	Appointed to editorial board
Michael Bowling	Robotics: Science and Systems Conference, 2006	Program committee
Michael Bowling	5th International Joint Conference on Autonomous Agents and Multi-Agent System	Senior program committee
Dale Schuurmans	23rd International Conference on Machine Learning	Senior program committee
Dale Schuurmans	17th European Conference on Machine Learning and the Tenth European Conference on Principles and Practice of Knowledge Discovery in Databases	Senior program committee
Dale Schuurmans	Journal of *Machine Learning Research*	Action editor

Dale Schuurmans	*Machine Learning* Journal	Associate editor
Vadim Bulitko	Workshop on "Planning and Learning in A Priori Unknown or Dynamic Domains" at the 19th International Joint Conference on Artificial Intelligence	Co-chair
Vadim Bulitko	Workshop on "Planning, Learning and Monitoring with Uncertainty and Dynamic Worlds" at the 17th biennial European Conference on Artificial Intelligence (ECAI-2006)	Organizing committee
Vadim Bulitko	19th International Joint Conference on Artificial Intelligence (IJCAI 2005)	Program committee
Vadim Bulitko	21st National Conference on Artificial Intelligence (AAAI 2006)	Program committee
Mohammad Ghavamzadeh	23rd International Conference on Machine Learning	Program committee
Mohammad Ghavamzadeh	5th International Joint Conference on Autonomous Agents and Multiagent Systems	Program committee
Yaakov Engel	22nd International Conference on Machine Learning	Program committee
Yaakov Engel	23rd International Conference on Machine Learning	Program committee
Yaakov Engel	20th Annual Conference on Neural Information Processing Systems	Program committee
Yaakov Engel	19th Annual Conference on Learning Theory	Program committee
Yaakov Engel	Workshop on "Kernel methods in Reinforcement Learning" at 23rd International Conference on Machine Learning	Program committee
Martin Zinkevich	5th International Joint Conference on Autonomous Agents and Multiagent Systems	Program committee
Martin Zinkevich	21st National Conference on Artificial Intelligence	Program committee
Martin Zinkevich	23rd International Conference on Machine Learning	Program committee
Martin Zinkevich	8th Workshop on "Game Theoretic and Decision Theoretic Agents" at 5th International Joint Conference on Autonomous Agents and Multiagent Systems	
Adam White	Workshop on "Reinforcement Learning: Benchmarks and Bake-offs II" at the 20th Annual Conference on Neural Information Processing Systems	Organizing committee
Brian Tanner	5th International Joint Conference on Autonomous Agents and Multiagent Systems.	Program committee

RLAI team members hosted many prominent short-term visitors during the reporting period, including: Yew Jin Lim (National University of Singapore), Doina Precup (McGill University), Michael James (Toyota Research), Sham Kakade (Toyota Technology Institute), Csaba Szepesvari (Computer and Automation Research Institute of the Hungarian Academy of Sciences), Khashayar Rohanimanesh (University of Massachussetts), Dengyong Zhou (NEC Laboratories America), Brett Browning (Carnegie Mellon University), Pascal Poupart (University of Waterloo).

GRADUATES

Student	Supervisor	Degree/Research topic	Currently
Peter McCracken	Bowling	MSc/Predictive state representations	IBM Research
Brian Tanner	Sutton	MSc/Temporal difference networks	Phd student at UofA
Greg Lee	Bulitko	MSc	Phd student at UofA
Finnegan Southey	Schuurmans	Phd	Google
Ali Ghodsi	Schuurmans/Frey	Phd	Asst. professor at the University of Waterloo

INTELLECTUAL PROPERTY

None.

PUBLICATIONS

REFEREED JOURNAL PUBLICATIONS

T. Caetano, T. Caelli, D.,Schuurmans, and D. Barrone, "Graphical Models and Point Pattern Matching," to appear in *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

C. Boutilier, R. Patrascu, P. Poupart, and D. Schuurmans, "Constraint-based Optimization and Utility Elicitation Using the Minimax Decision Criterion," to appear in *Artificial Intelligence*.

V. Bulitko and G. Lee, "Learning in Real Time Search: A Unifying Framework," *Journal of Artificial Intelligence Research*, 2006, pp. 25:119-157.

Y. Yekutieli, R. Sagiv-Zohar, R. Aharonov, Y. Engel, B. Hochner and T. Flash, "A Dynamic Model of the Octopus Arm. I. Biomechanics of the Octopus Reaching Movement," *Journal of Neurophysiology*, 2005, pp. 94: 1443-1458.

M. Ghavamzadeh, S. Mahadevan, and R. Makar, "Hierarchical Multiagent Reinforcement Learning," *Journal of Autonomous Agents and Multi-Agent Systems*, 2006.

REFERREED CONFERENCE PROCEEDINGS

D. Precup, R.S. Sutton, C. Paduraru, A. Koop, and S. Singh, "Off-policy Learning with Recognizers," in *Advances in Neural Information Processing Systems 18* (NIPS-05), 2005. 27% acceptance rate.

R.S. Sutton, E. Rafols, and A. Koop, "Temporal Abstraction in Temporal-difference Networks," in *Advances in Neural Information Processing Systems 18* (NIPS-05), 2005. 27% acceptance rate.

E.J. Rafols, M.B Ring, R.S. Sutton, and B. Tanner, "Using Predictive Representations to Improve Generalization in Reinforcement Learning," in *Proc. 19th Int'l Joint Conf. on Artificial Intelligence*, 2005, pp. 835-840. 18% acceptance Rate.

Geramifard, M. Bowling, R. Sutton, "Incremental Least-Square Temporal Difference Learning," to appear in *Proceedings of the 21st National Conf. on Artificial Intelligence* (AAAI-06), 2006. 30% acceptance rate.

L. Xu, and D. Schuurmans, "Discriminative Unsupervised Learning of Structured Predictors," to appear in *Proc. of the 23rd Int'l Conf Machine Learning* (ICML-06), Pittsburgh, Penn, June 25-29, 2006. 20% acceptance rate.

L. Xu, K. Crammer, and D. Schuurmans, "Robust Support Vector Machine Training via Convex Outlier Ablation," to appear in *Proc. of the 21st National Conf. on Artificial Intelligence* (AAAI-06), 2006. 22% acceptance rate.

Y. Guo, D. Wilkinson, and D. Schuurmans, "Maximum Margin Bayesian Networks," *Proc. of the Conf. on Uncertainty in Artificial Intelligence* (UAI-05), 2005. 34% acceptance rate.

Q. Wang, D. Schuurmans, and D. Lin, "Strictly Lexical Parsing," *Proc. of the IEEE Int'l Conf. on Natural Language Processing and Knowledge Engineering*, 2005. 55% acceptance rate.

T. Wang, M. Bowling, P. Poupart, and D. Schuurmans, "Compact, Convex Upper Bound Iteration for Approximate POMDP Planning," to appear in *Proceedings of the 21st National Conf. on Artificial Intelligence* (AAAI-06), 2006. 22% acceptance rate.

Q. Wang, C. Cherry, D. Lizotte, and D. Schuurmans, "Improved large margin dependency parsing via local constraints and Laplacian regularization," to appear in *Proceedings of the 10th Conf. on Computational Natural Language Learning* (CONLL-06), 2006.

F. Southey, M. Bowling, B. Larson, C. Piccione, N. Burch, and D. Billings, "Bayes' Bluff: Opponent Modeling in Poker," to appear in the *Proceedings of the Conf. on Uncertainty in Artificial Intelligence* (UAI-06), Cambridge, MA, USA, July 13-16, 2006. 34% acceptance rate.

M. Bowling, D. Wilkinson, and A. Milstein, "Subjective Localization with Action Respecting Embedding," in *Proc. of the 12th International Symposium of Robotics Research* (ISRR-05), San Francisco, CA, USA, October 12-15, 2005. 30% acceptance rate.

P. McCracken and M. Bowling, "Online Discovery and Learning of Predictive State Representations," in *Advances in Neural Information Processing Systems 18* (NIPS-05), 2005. 27% acceptance rate.

N.Sturtevant and M. Bowling, "Robust Game Play Against Unknown Opponents," *5th Int'l Joint Conf. on Autonomous Agents and Multiagent Systems* (AAMAS-06), Hakodate, Japan, May 8-12, 2006. 23% acceptance rate.

M. Bowling, P. McCracken, M. James, J. Neufeld, D. Wilkinson. "Learning Predictive State Representations Using Non-Blind Policies," to appear in *Proc. of the 23rd Int'l Conf. on Machine Learning* (ICML-06), 2006. 20% acceptance rate.

A. Rettinger, M. Zinkevich, M. Bowling, "Boosting Expert Ensembles for Rapid Concept Recall," to appear in *Proc. of the 21st National Conf. on Artificial Intelligence* (AAAI), 2006. 30% acceptance rate.

A. Kaboli, M. Bowling, P. Musilek, "Bayesian Calibration for Monte Carlo Localization," to appear in the *Proc. of the 21st National Conference on Artificial Intelligence* (AAAI-06), 2006. 30% acceptance rate.

N. Sturtevant, M.Zinkevich, and M.Bowling, "ProbMaxn: Opponent Modeling in N-Player Games," to appear in *Proc. of the 21st National Conf. on Artificial Intelligence* (AAAI-06), 2006. 30% acceptance rate.

M. Zinkevich, M. Bowling, N. Bard, M. Kan, and D. Billings, "Optimal Unbiased Estimators for Evaluating Agent Performance," to appear in the *Proc. of the 21st National Conf on Artificial Intelligence* (AAAI), 2006. 30% acceptance rate.

M. Bowling, D. Wilkinson, and A. Ghodsi, "Subjective Mapping," to appear in the New Scientific and Technical Advances in Research (NECTAR) of the *Proc. of the 21st National Conference on Artificial Intelligence*, 2006. 55% acceptance rate.

M. Zinkevich, A. Greenwald, and M. Littman, "Cyclic Equilibria in Markov Games," *Advances in Neural Information Processing Systems 18* (NIPS-05), 2005. 27% acceptance rate.

V.Bulitko, N. Sturtevant and M. Kazakevich, "Speeding Up Learning in Real-time Search via Automatic State Abstraction," in *Proc. of the National Conf. on Artificial Intelligence* (AAAI-05), 2005, pp. 1349-1354.

G.Lee and V.Bulitko, "Genetic Algrorithms for Action Set Selection Across Domains: A Demonstration," in *Proc. of the Genetic and Evolutionary Computation Conference* (GECCO), 2006.

G. Lee and V. Bulitko, "GAMM: Genetic Algorithms with Meta-Models for Vision," in *Proc. of the Genetic and Evolutionary Computation Conf.* (GECCO), 2005, pp. 2029-2036.

V. Bulitko and D. Wilkins, "Machine Learning for Time Interval Petri Nets," in Lecture Notes in Artificial Intelligence (LNAI), *Proc. of the 18th Australian Joint Conf. on Artificial Intelligence*, 2005, pp. 959 - 965.

D. Thue and V.Bulitko, "Modelling Goal-directed Players in Digital Games," in *Proc. of the Artificial Intelligence and Interactive Digital Entertainment Conf.* (AIIDE-06), 2006.

N. Sturtevant, V. Bulitko and M. Buro, "Automatic State Abstraction for Pathfinding in Real-Time Video Games," in *Proceedings of the Int'l Symposium on Abstraction, Reformulation and Approximation* (SARA), 2005, pp. 362 - 364.

Y. Engel, P. Szabo and D. Volkinshtein, "Learning to Control an Octopus Arm with Gaussian Process Temporal Difference Methods," in *Advances in Neural Information Processing Systems 18* (NIPS), 2005. 27% acceptance rate.

Y. Engel, S. Mannor, R. Meir, "Reinforcement Learning with Gaussian Processes," in *International Conference on Machine Learning* (ICML-05), 2005. 30% acceptance rate.

N.R. Sturtevant, and A.M. White, "Feature Construction for Reinforcement Learning in Hearts," in *Proc. 5th International Conference on Computers and Games*, 2006.

N. Aharony, T. Zehavi, and Y. Engel, "Learning Wireless Network Association-Control with Gaussian Process Temporal Difference Methods," *Proc. OPNETWORK 2005*. (Best Technical Paper Award winner).

BOOKS AND CHAPTERS

D. Schuurmans, F. Southey, D. Wilkinson, and Y. Guo, "Metric-based Approaches for Semi-supervised Regression and Classification," To appear in "*Semi-Supervised Learning,*" edited by O. Chapelle, B. Schoelkopf and A. Zein, MIT Press.

Levner, V. Bulitko and G. Lin, "Feature Extraction for Classification of Proteomic Mass Spectra: A Comparative Study," *in Feature Extraction, Foundations and Applications*, 2005, Springer.

D. Silver, "Cooperative Pathfinding," in *AI Game Programming Wisdom 3*, 2006, Charles River Media, pages 99-111.

SPECIAL INVITED PRESENTATIONS

Richard Sutton, "Experience-Oriented Artificial Intelligence," distinguished lecture series of the Department of Mathematics and Computer Science at Lethbridge University, April, 2005.

Richard Sutton, "Experience-Oriented Artificial Intelligence," distinguished lecture series at the Cognitive Science Department of the University of California at San Diego, May, 2005.

Richard Sutton, "Grounding Knowledge in Subjective Experience," invited talk at the Second Cognitive Systems Conference, Arlington, VA, May, 2005.

Richard Sutton "Predictive Representations of State and Knowledge," invited talk at the workshop on "Rich representations for reinforcement learning" at the 2005 International Conference on Machine Learning, Bonn, Germany, August, 2005.

Richard Sutton, "Reinforcement Learning and Animal Behavior," invited talk at the Workshop on Dopamine and its Role in Learning, Motivation, and Psychiatric Disorders, Douglas Research Centre, McGill University, December, 2005.

Richard Sutton, "Carving the World at its Joints," invited talk at the NIPS-2005 Workshop on Towards Human-level Artificial Intelligence, Whistler, BC, December, 2005.

Michael Bowling, "Bayes' Bluff: Opponent Modeling in Poker," a demonstration at the International Conference on Neural Information Processing Systems, Vancouver, BC, December, 2005.

AWARDS / THESES

Distinguished Paper Award, International Joint Conference on Artificial Intelligence (IJCAI) 2005, Y. Guo, R. Greiner, and D. Schuurmans, "Learning Coordination Classifiers."

M. Bowling - Finalist for the Alberta Science and Technology (ASTech) "Leaders of Tomorrow" award.

Schuurmans - Research Award, Department of Computing Science, University of Alberta

Schuurmans - Faculty of Science nominee for "Martha Cook Piper Research Prize."

R. Sutton, D. Schuurmans, and M. Bowling – Finalists as principal investigators of the Alberta Ingenuity Centre for Machine Learning for the Alberta Science and Technology (ASTech) "Outstanding Leadership in Alberta Technology" award.

Best Technical Paper Award, Opnetwork 2005, N. Aharony, T. Zehavi, and Y. Engel, "Learning Wireless Network Association-Control with Gaussian Process Temporal Difference Methods."

Bowling accepted an appointment as an Adjunct Assistant Professor in the School of Computer Science at the University of Waterloo.

OUTREACH

Press:

Edmontonians (cover article), November 2005, Vol. XVI No. 11
Edmonton Journal, December 14, 2005
Daily Planet, January 24, 2006

High School Presentations:

"Games Robots Play," Meridian Heights School, Stony Plain, May 2005 – Michael Bowling.

"Games Robots Play," and "Start Now," Alliance High School, Alliance, Ohio, May, 2005 – Michael Bowling.

Public Presentation:

"Robots that Learn," Odyssium, Edmonton, April, 2005 - Richard Sutton

Undergraduate Lectures:

"Robots that Learn," Lethbridge University, April, 2005 - Richard Sutton

"Robots that Learn," McGill University, November, 2005 - Richard Sutton

WISEST summer student:

Amanda Brewer, June – July 2005

FUNDING

Summary of funding from other sources:

Canada research chair to Schuurmans	100,000
NSERC discovery grant to Sutton	50,000
NSERC discovery grant to Schuurmans	47,000
NSERC discovery grant to Bowling	26,000
NSERC discovery grant to Bulitko	15,000
DARPA grant to Universities of Rutgers, Michigan, and Massachusetts; supports Mark Ring	~480,000/yr, RLAI portion 36,000
AIF postdoctoral fellowship to Yaakov Engel	110,000
MITACS grant "Statistical Learning of Complex Data with Complex Distributions" to Schuurmans, Sutton, with Yoshua Bengio, Hugh Chipman, Christian Leger, William Welch, Jim Ramsay, and Mu Zhu	170,000/yr, RLAI portion ~25,000
The Alberta Ingenuity Center for Machine Learning (AICML), PIs Sutton, Bowling, Schuurmans, Holte, Greiner, Schaeffer, and Goebel, provided support for programming staff, software management, visitors, robot lab renovation, robot hardware, software support, postdocs and students	~2M/yr, RLAI portion ~800,000
Scholarships: AIF, NSERC, iCORE, University, Killam, Steinhauer, Sampson, Ontario, Reporting period only.	~367,000
NSERC Strategic Grant, "Intelligent Agents for Interactive Entertainment" to Bowling, Schuurmans, with Jonathan Schaeffer, Robert Holte, Duane Szafron, and Michael Buro	195,000/yr, RLAI portion ~30,000
Advanced International Education and Research Support Program - Oita University Supports Katsunari Shibata	46,600
Electronic Arts/NSERC to John Buchanan; provided summer support to Michael Johanson (MSc student)	~8,000